

Deblending of Simultaneous-source Seismic Data using Fast Iterative Shrinkage-thresholding Algorithm with Firm-thresholding

Shan QU¹, Hui ZHOU¹, Renwu LIU², Yangkang CHEN³, Shaohuan ZU¹, Sa YU¹, Jiang YUAN¹, and Yahui YANG¹

¹State Key Laboratory of Petroleum Resources and Prospecting,
Key Lab of Geophysical Exploration of CNPC, Dept. of Geophysics,
China University of Petroleum, Beijing, China;
e-mail address: huizhou@cup.edu.cn (corresponding author)

²Dagang Geophysical Prospecting Branch of BGP, CNPC, Tianjin, China

³The University of Texas at Austin, Austin, USA

Abstract

In this paper, an improved algorithm is proposed to separate blended seismic data. We formulate the deblending problem as a regularization problem in both common receiver domain and frequency domain. It is suitable for different kinds of coding methods such as random time delay discussed in this paper. Two basic approximation frames, which are iterative shrinkage-thresholding algorithm (ISTA) and fast iterative shrinkage-thresholding algorithm (FISTA), are compared. We also derive the Lipschitz constant used in approximation frames. In order to achieve a faster convergence and higher accuracy, we propose to use firm-thresholding function as the thresholding function in ISTA and FISTA. Two synthetic blended examples demonstrate that the performances of four kinds of algorithms (ISTA with soft- and firm-thresholding, FISTA with soft- and firm-thresholding) are all effective, and furthermore FISTA with a firm-thresholding operator exhibits the most robust

behavior. Finally, we show one numerically blended field data example processed by FISTA with firm-thresholding function.

Key words: deblending, regularization, simultaneous-source, firm-thresholding.

1. INTRODUCTION

The simultaneous-source technique is an efficient seismic acquisition technology; however, blended seismic data acquired by this technology should be separated before they are processed using the processing methods developed for conventional unblended data. Because of both its economic benefits and technical challenges, this technique has attracted the attention of researchers in both industry and academia (Mahdad *et al.* 2011, Abma *et al.* 2010, Huo *et al.* 2009, Chen *et al.* 2014a).

Different filtering and inversion methods have been applied previously to deblend seismic data. Deblending processing can be considered as a denoising problem because the blending noise appears as random noise when the blended data are sorted into a different domain, *e.g.*, common receiver, common offset or common midpoint domain. According to this property, Beasley (2008) used a geometry-related filter to separate marine blended data in the common midpoint domain. Huo *et al.* (2009) introduced a multi-directional vector median filter after resorting the data into common midpoint gathers. Chen *et al.* (2014b) proposed a deblending method by using normal moveout and median filtering in common-midpoint gathers. However, the denoising methods do not work well when the blended data are complex. Deblending problem can also be formulated as an inversion problem which aims at estimating the desired unknown unblended data. Because of the ill-posedness of such estimation problems, sophisticated regularization plays an important role in the deblending procedure (Doulgeris *et al.* 2012). Akerberg *et al.* (2008) used a sparse constraint to regularize the ill-posed problem in radon domain. Lin and Herrmann (2009) transformed the data into curvelet domain, where the data also present the sparsity, to regularize the inversion. Mahdad *et al.* (2011, 2012), Doulgeris *et al.* (2012), and Chen and Ma (2014) introduced an iterative f-k filtering method that combines the properties of filtering and inversion methods in different domains, but only when the number of simultaneous-source blended shots is greater than or equal to 3, these methods are effective. Chen *et al.* (2014a) proposed a novel iterative estimation scheme using shaping regularization in the seislet domain, but the blending matrix in this method is only a concept matrix, which limits the application of different iterative estimation schemes. In addition, none of the existing methods is suitable for different coding method, such as phase coding or sweep length.

In this paper, an improved algorithm is introduced to separate the blended seismic data. First of all, we define the deblending problem as a regularization problem in both common receiver domain and frequency domain. Then, we compare two basic approximation frames: ISTA and FISTA, and derive the Lipschitz constant. It has been proved that FISTA has a faster convergence speed than ISTA. After that, we propose to use firm-thresholding function as the thresholding function in ISTA and FISTA in order to achieve a higher accuracy. We apply those algorithms to remove blending noise, and meanwhile to preserve useful components of seismic data.

This paper is organized as follows. In Section 2, we present the definition of the deblending problem. In Section 3, we first introduce two basic approximation frames (ISTA and FISTA) and compare their convergence rate. Then, we discuss the reduction of Lipschitz constant used in the approximation frames. Next, we compare the traditional soft-thresholding function and hard-thresholding function with a better choice firm-thresholding. In Section 4, we test the proposed algorithms on two examples of numerically blended synthetic data and one example of numerically blended field data. The algorithms, which are ISTA with soft-thresholding, ISTA with firm-thresholding, FISTA with soft-thresholding, and FISTA with firm-thresholding, are compared before some conclusions are given.

2. THE DEFINITION OF DEBLENDING PROBLEM

This deblending model can be applied to any number of sources, while in this paper we only discuss the case of two independent sources. Supposing there are two independent sources corresponding to two shooting vessels in the Ocean Bottom Cable (OBC) acquisition. Here, one source means a collection of shots from one shooting vessel. The two sources are fired with small random time delay. The acquisition geometry is shown in Fig. 1.

The forward problem can be formulated as

$$\mathbf{d} = \mathbf{d}_1 + \mathbf{T}\mathbf{d}_2, \quad (1)$$

where \mathbf{d} denotes the blended data, \mathbf{d}_1 and \mathbf{d}_2 are unblended seismic datasets excited, respectively, by the two sources, and \mathbf{T} stands for the blending operator in the time domain.

Chen *et al.* (2014a) proposed to augment Eq. 1 with another equation through the inverse blending operator \mathbf{T}^{-1}

$$\mathbf{T}^{-1}\mathbf{d} = \mathbf{T}^{-1}\mathbf{d}_1 + \mathbf{d}_2, \quad (2)$$

then the blending problem can be formulated in the common receiver gather and time domain as

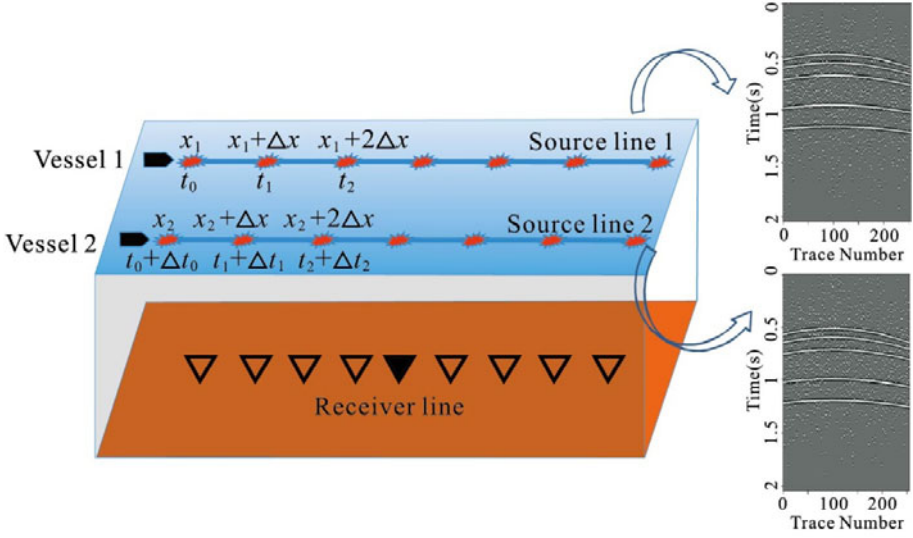


Fig. 1. The illustration of the simultaneous-source technique corresponding to two shooting vessels on the ocean surface in the OBC acquisition. ∇ and \blacktriangledown denote, respectively, receivers and a common receiver used to generate the blended data shown on the right side, and \star represents shot positions. x_1 and x_2 stands for the first shot position in source lines 1 and 2, respectively, Δx is the shot point interval, t_k denotes the shooting time of the k -th shot on source line 1, Δt_k represents time delay of the k -th shot on source line 2. Two profiles demonstrate the blending seismic datasets in the common receiver domain corresponding to source lines 1 and 2, respectively.

$$\tilde{\mathbf{F}}\tilde{\mathbf{m}} = \tilde{\mathbf{d}}, \quad (3)$$

where $\tilde{\mathbf{d}} = \begin{bmatrix} \mathbf{d} \\ \mathbf{T}^{-1}\mathbf{d} \end{bmatrix}$, $\tilde{\mathbf{F}} = \begin{bmatrix} \mathbf{I} & \mathbf{T} \\ \mathbf{T}^{-1} & \mathbf{I} \end{bmatrix}$, $\tilde{\mathbf{m}} = \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix}$, \mathbf{I} is a unit matrix.

In addition, instead of solving the original forward problem, Chen *et al.* (2014a) turn to solve the augmented Eq. 3, because two datasets can be obtained simultaneously, although this augmented equation is mathematically as ill-posed as the original forward problem. Deblending problem is not like noise attenuation, where the noise part is treated as useless and should be removed. The augmented equation also allows some basic frameworks, which will be discussed in Section 3.1.

Besides, the operator \mathbf{T} in Eq. 3 has the following form

$$\mathbf{T} = \mathcal{F}^{-1}\mathbf{P}\mathcal{F}, \quad (4)$$

where \mathcal{F} and \mathcal{F}^{-1} are forward and inverse Fourier transforms, respectively, and \mathbf{P} is a block phase-shift operator, which can be expressed as

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & & & \\ & \mathbf{P}_2 & & \\ & & \ddots & \\ & & & \mathbf{P}_N \end{bmatrix}, \quad (5)$$

where diagonal matrix $\mathbf{P}_n (n = 1, 2, \dots, N)$ denotes the individual phase shift operator for the n -th trace having the following form

$$\mathbf{P}_n = \text{diag} \left(e^{-2\pi i f \Delta t_n} \quad e^{-2\pi i f \Delta t_n} \quad \dots \quad e^{-2\pi i f \Delta t_n} \right)_{M \times M}, \quad (6)$$

where Δt_n means the random time delay of the n -th trace, and M denotes the number of temporal samples.

According to the representation of \mathbf{T} , we can see that it is a conceptual blending operator, and it cannot be formulated by a specific matrix. In order to overcome this disadvantage, we transform the forward problem (Eq. 1) into the frequency domain

$$\mathbf{D} = \mathbf{D}_1 + \mathbf{\Gamma} \mathbf{D}_2, \quad (7)$$

where \mathbf{D} denotes the blended data in the frequency domain, \mathbf{D}_1 and \mathbf{D}_2 represent the frequency domain data of \mathbf{d}_1 and \mathbf{d}_2 , and $\mathbf{\Gamma}$ is the blending operator in the frequency domain.

Here, $\mathbf{\Gamma}$ is not just a simple Fourier Transform of the blending operator \mathbf{T} in Eq. 4. For a given frequency, it is a specific matrix. Considering one receiver gather and random time delay coding, $\mathbf{\Gamma}$ at frequency f can be formulated by

$$\mathbf{\Gamma} = \text{diag} \left(e^{-2\pi i f \Delta t_1} \quad e^{-2\pi i f \Delta t_2} \quad \dots \quad e^{-2\pi i f \Delta t_N} \right)_{N \times N}, \quad (8)$$

where $\Delta t_n (n = 1, 2, \dots, N)$ denotes the random time delay of the n -th shot, and N stands for the number of shots.

Then, by the same token as Chen *et al.* (2014a), we augment Eq. 7 into the following form

$$\mathbf{F} \mathbf{m} = \tilde{\mathbf{D}}, \quad (9)$$

where $\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} \\ \mathbf{\Gamma}^{-1} \mathbf{D} \end{bmatrix}$, $\mathbf{F} = \begin{bmatrix} \mathbf{I} & \mathbf{\Gamma} \\ \mathbf{\Gamma}^{-1} & \mathbf{I} \end{bmatrix}$, $\mathbf{m} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix}$.

There are two main reasons for us to deblend the data in the frequency domain. The first reason is that unlike blending operator described in Eqs. 4-

6, the new blending operator in the frequency domain in Eq. 8 can be formulated by a specific matrix for one frequency. This property turns the deblending problem to a traditional regularization problem for each frequency and allows more approximation frames to be discussed in Section 3. The second reason is that the new blending operator allows more kinds of coding methods, such as phase coding and sweep length coding in the vibroseis acquisition (Qu *et al.* 2014). As a tutorial, Γ is

$$\Gamma = \text{diag}\left(e^{i\beta_1} \quad e^{i\beta_2} \quad \dots \quad e^{i\beta_N}\right)_{N \times N}, \quad (10)$$

for phase coding, $\beta_n (n = 1, 2, \dots, N)$ represents the phase code of the n -th shot, and Γ is written as

$$\Gamma = \text{diag}\left(\frac{1}{\sqrt{q_1}} e^{\left[\frac{q_1-1}{q_1} \pi i \frac{(f-f_0)^2}{2k}\right]} \quad \frac{1}{\sqrt{q_2}} e^{\left[\frac{q_2-1}{q_2} \pi i \frac{(f-f_0)^2}{2k}\right]} \quad \dots \quad \frac{1}{\sqrt{q_N}} e^{\left[\frac{q_N-1}{q_N} \pi i \frac{(f-f_0)^2}{2k}\right]}\right)_{N \times N}, \quad (11)$$

for sweep length coding, where k represents the sweep rate and $q_n (n = 1, 2, \dots, N)$ represents that the sweep rate of the encoded signal of the n -th shot is q_n times as that of the original signal.

It is known that seismic data often has sparse representation in some transformation domains, *e.g.*, curvelet domain (Candes and Demanet 2005). One way to overcome the ill-posedness of Eq. 9 is to minimize an objective function $J(\mathbf{m})$:

$$\begin{aligned} J(\mathbf{m}) &= f(\mathbf{m}) + g(\mathbf{m}), \\ f(\mathbf{m}) &= \|\tilde{\mathbf{D}} - \mathbf{F}\mathbf{m}\|_2^2, \\ g(\mathbf{m}) &= \lambda \|\mathbf{C}^{-1}\mathbf{m}\|_p \quad (0 \leq p \leq 1), \end{aligned} \quad (12)$$

where \mathbf{C}^{-1} is sparsity-promoting transform that maps the signal to the transformation domain, such as curvelet domain, $f(\mathbf{m})$ is a data-fitting term and $g(\mathbf{m})$ is a regularization term used to penalize the complex model, and λ is the penalty term.

3. DEBLENDING BY USING IMPROVED ALGORITHM

3.1 The basic approximation frames and their convergence rates

Several famous iterative approximation frames for solving regularization problem have been fully developed and applied to image denoising and restoration problems. The optimization problem like Eq. 9 belongs to regularization problem. For the purpose of comparison, we focus on the following two approximation frames.

3.1.1 Iterative shrinkage-thresholding algorithm (ISTA)

The most popular method to solve a regularization problem is ISTA because of its convenient implementation. ISTA was proposed by Daubechies *et al.* (2004), Figueiredo and Nowak (2003), and Chambolle (2004) to solve least-square minimization problem with a constraint to the model. We modify the basic form of ISTA to correspond to the deblending problem as

$$\mathbf{m}_{k+1} = \mathbf{C}T_{\tau,p}\mathbf{C}^{-1}\left[\mathbf{m}_k - \frac{2}{L(\nabla f(\mathbf{m}))}\mathbf{F}^T(\mathbf{F}\mathbf{m}_k - \tilde{\mathbf{D}})\right], \quad (13)$$

where \mathbf{m}_k represents the solution at the k -th iteration, $L(\nabla f(\mathbf{m}))$ is the Lipschitz constant of $\nabla f(\mathbf{m})$ depending on the maximum eigenvalue of $\mathbf{F}^T\mathbf{F}$. This constant is not as easy to calculate as the traditional inverse problem, because \mathbf{F} depends on the frequency as we discussed in Section 2. We will discuss it in Section 3.2. $T_{\tau,p}$ is a thresholding operator corresponding to specific norm regularization to be shown in Section 3.3.

The convergence rate of ISTA has been well investigated for the regularization problem. It is $O(1/k)$, which is also called a “sublinear global rate of convergence”. That means each value of the sequence $\{J(\mathbf{m}_k)\}$ converges to the optimal function value $J(\mathbf{m}_*)$, and it is proved to behave like

$$J(\mathbf{m}_k) - J(\mathbf{m}_*) \leq \frac{2L(\nabla f(\mathbf{m}))\|\mathbf{m}_0 - \mathbf{m}_*\|^2}{k}, \quad (14)$$

thus

$$J(\mathbf{m}_k) - J(\mathbf{m}_*) \leq O(1/k), \quad (15)$$

where \mathbf{m}_* is the optimal solution, and \mathbf{m}_0 is an initial model.

3.1.2 Fast iterative shrinkage-thresholding algorithm (FISTA)

FISTA proposed by Beck and Teboulle (2009) is a variation of ISTA. We modify the basic form of FISTA to be suitable for the deblending problem as

$$\mathbf{m}_{k+1} = \mathbf{W}\left(\mathbf{m}_k + \frac{t_k - 1}{t_{k+1}}(\mathbf{m}_k - \mathbf{m}_{k-1})\right), \quad (16)$$

where $t_{k+1} = \left(1 + \sqrt{1 + 4(t_k)^2}\right)/2$, $t_0 = 1$, and the operator \mathbf{W} is given by

$$\mathbf{W}(\mathbf{m}) = \mathbf{C}T_{\tau,p}\mathbf{C}^{-1}\left[\mathbf{m} - \frac{2}{L(\nabla f(\mathbf{m}))}\mathbf{F}^T(\mathbf{F}\mathbf{m} - \tilde{\mathbf{D}})\right]. \quad (17)$$

FISTA seems to have the same complexity as the ISTA, but it has been proved to have a faster convergence rate. The sequence $\{\mathbf{m}_k\}$ generated by FISTA achieves

$$J(\mathbf{m}_k) - J(\mathbf{m}_*) \leq \frac{2L(\nabla f(\mathbf{m})) \|\mathbf{m}_0 - \mathbf{m}_*\|^2}{(k+1)^2}, \quad (18)$$

and thus

$$J(\mathbf{m}_k) - J(\mathbf{m}_*) \leq O(1/k^2). \quad (19)$$

3.2 The reduction of Lipschitz constant

It is noted from Eq. 8 that $\mathbf{\Gamma}$ is a function of frequency, and that it is also different from different coding methods as shown in Eqs. 10 and 11. In this part, we prove that $L(\nabla f(\mathbf{x}))$ can be formulated by a constant.

It can be derived from Eq. 8 that $\mathbf{\Gamma}^{-1} = \mathbf{\Gamma}^T$, where $\mathbf{\Gamma}^T$ means the conjugate transpose of $\mathbf{\Gamma}$, then

$$\mathbf{F}^T \mathbf{F} = \begin{bmatrix} \mathbf{I} & \mathbf{\Gamma} \\ \mathbf{\Gamma}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{\Gamma} \\ \mathbf{\Gamma}^{-1} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} 2\mathbf{I} & 2\mathbf{\Gamma} \\ 2\mathbf{\Gamma}^{-1} & 2\mathbf{I} \end{bmatrix}. \quad (20)$$

According to the linear algebra knowledge, the eigenvalues λ of $\mathbf{F}^T \mathbf{F}$ satisfy

$$|\mathbf{F}^T \mathbf{F} - \lambda \mathbf{I}| = 0, \quad (21)$$

and by using Eq. 20 we have

$$|\mathbf{F}^T \mathbf{F} - \lambda \mathbf{I}| = \begin{vmatrix} (2-\lambda)\mathbf{I} & 2\mathbf{\Gamma} \\ 2\mathbf{\Gamma}^{-1} & (2-\lambda)\mathbf{I} \end{vmatrix}. \quad (22)$$

According to the block matrix property, Eq. 22 can be modified to

$$|\mathbf{F}^T \mathbf{F} - \lambda \mathbf{I}| = |\lambda(\lambda - 4)| = 0. \quad (23)$$

Then we obtain that the maximum eigenvalue of $\mathbf{F}^T \mathbf{F}$ is 4, and as a result, $L(\nabla f(\mathbf{m})) = 2\lambda_{\max}(\mathbf{F}^T \mathbf{F}) = 8$ for each frequency.

The whole reduction of Lipschitz constant is also suitable for higher number of blended sources. Now we will try to derive the Lipschitz constant when the number of blended sources is K . The blending equation in the frequency domain is as follows

$$\mathbf{F}\mathbf{m} = \tilde{\mathbf{D}}, \quad (24)$$

$$\text{where } \tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} \\ \Gamma_2^{-1} \mathbf{D}_2 \\ \vdots \\ \Gamma_K^{-1} \mathbf{D}_K \end{bmatrix}, \mathbf{F} = \begin{bmatrix} \mathbf{I} & \Gamma_2 & \cdots & \Gamma_K \\ \Gamma_2^{-1} & \mathbf{I} & \cdots & \Gamma_2^{-1} \Gamma_K \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_K^{-1} & \Gamma_K^{-1} \Gamma_2 & \cdots & \mathbf{I} \end{bmatrix}, \mathbf{m} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_K \end{bmatrix},$$

$\mathbf{D}_k (k = 2, \dots, K)$ is the k -th source, $\Gamma_k (k = 2, \dots, K)$ is the frequency blending operator for the k -th source.

Then,

$$\mathbf{F}^T \mathbf{F} = \begin{bmatrix} K\mathbf{I} & K\Gamma_2 & \cdots & K\Gamma_K \\ K\Gamma_2^{-1} & K\mathbf{I} & \cdots & K\Gamma_2^{-1}\Gamma_K \\ \vdots & \vdots & \ddots & \vdots \\ K\Gamma_K^{-1} & K\Gamma_K^{-1}\Gamma_2 & \cdots & K\mathbf{I} \end{bmatrix}, \quad (25)$$

$$|\mathbf{F}^T \mathbf{F} - \lambda \mathbf{I}| = \begin{vmatrix} (K - \lambda)\mathbf{I} & K\Gamma_2 & \cdots & K\Gamma_K \\ K\Gamma_2^{-1} & (K - \lambda)\mathbf{I} & \cdots & K\Gamma_2^{-1}\Gamma_K \\ \vdots & \vdots & \ddots & \vdots \\ K\Gamma_K^{-1} & K\Gamma_K^{-1}\Gamma_2 & \cdots & (K - \lambda)\mathbf{I} \end{vmatrix}. \quad (26)$$

According to the linear algebra knowledge, by using some elementary transformations, Eq. 26 can be simplified as

$$|\mathbf{F}^T \mathbf{F} - \lambda \mathbf{I}| = \begin{vmatrix} (2K - \lambda)\mathbf{I} & K\Gamma_2 & \cdots & K\Gamma_K \\ 0 & -\lambda\mathbf{I} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\lambda\mathbf{I} \end{vmatrix} = |(\lambda - 2K)\lambda^{K-1}|. \quad (27)$$

Now we obtain that the maximum eigenvalue of $\mathbf{F}^T \mathbf{F}$ is $2K$. As a result, $L(\nabla f(\mathbf{m})) = 2\lambda_{\max}(\mathbf{F}^T \mathbf{F}) = 4K$ for each frequency. It can be seen that when $K = 2$, $L(\nabla f(\mathbf{m})) = 8$. This result matches the previous conclusion.

Here is an important part we need to discuss. In terms of complex real simultaneous-source seismic data, $L(\nabla f(\mathbf{m}))$ should be set slightly larger than $4K$, because some parts of the complex blended profiles are actually blended by larger than $4K$ sources.

3.3 Thresholding functions

As we discussed in Section 2, in order to overcome the ill-posedness of Eq. 9, we need to recover the single seismic data \mathbf{m} from the blended data $\tilde{\mathbf{D}}$, when \mathbf{m} is of the sparsest structure in a sparse domain; that means, \mathbf{m} has the fewest nonzero components in the sparse domain. Accordingly, the sparsest possible representation of the debrending problem can be modeled as the

minimization of the following objective function $J(\mathbf{m})$ with L_0 norm regularization

$$J(\mathbf{m}) = \|\tilde{\mathbf{D}} - \mathbf{F}\mathbf{m}\|_2^2 + \lambda \|\mathbf{C}^{-1}\mathbf{m}\|_0, \quad (28)$$

where $\|\mathbf{C}^{-1}\mathbf{m}\|_0$ denotes L_0 norm regularization.

Thresholding function is related to one specific norm regularization. When we use the objective function with L_0 norm regularization as shown in Eq. 28 ($p = 0$), $T_{\tau,0}$ in Eqs. 13 and 17 is a hard-thresholding operator with a thresholding value τ and takes the form of

$$T_{\tau,0} = \begin{cases} v(\mathbf{m}_k), & \text{for } |v(\mathbf{m}_k)| > \tau, \\ 0, & \text{for } |v(\mathbf{m}_k)| \leq \tau, \end{cases} \quad (29)$$

at the k -th iteration, $v(\mathbf{m}_k) = \mathbf{C}^{-1} \left[\mathbf{m}_k - \frac{2}{L(\nabla f(\mathbf{m}))} \mathbf{F}^T (\mathbf{F}\mathbf{m}_k - \tilde{\mathbf{D}}) \right]$.

Unfortunately, Bruce and Gao (1996) have shown that due to the discontinuity, the hard-thresholding ($T_{\tau,0}$) tends to have bigger variance, which would lead to unstable and non-convergent results. Moreover, Candes and Tao (2005) and Donoho (2006) have shown that solving Eq. 28 would appear to require combinatorial optimization, and be utterly intractable to solve. In other words, L_0 norm is the most accurate norm, because it conforms to the physical meaning. However, it is sometimes unstable and hard to solve. Therefore, they pointed that L_1 norm is in some sense the convex relaxation of the L_0 norm and it is more tractable than L_0 norm, L_1 norm can recover sparse unknown signals in a variety of different situations because of its stability. Equation 28 can be rewritten as

$$J(\mathbf{m}) = \|\tilde{\mathbf{D}} - \mathbf{F}\mathbf{m}\|_2^2 + \lambda \|\mathbf{C}^{-1}\mathbf{m}\|_1. \quad (30)$$

When we use the objective function with L_1 norm regularization ($p = 1$), $T_{\tau,1}$ is a soft-thresholding operator as the form of

$$T_{\tau,1} = \begin{cases} v(\mathbf{m}_k) - \tau \operatorname{sgn}(v(\mathbf{m}_k)), & \text{for } |v(\mathbf{m}_k)| > \tau, \\ 0, & \text{for } |v(\mathbf{m}_k)| \leq \tau, \end{cases} \quad (31)$$

where $\operatorname{sgn}(\cdot)$ represents sign function.

However, Bruce and Gao (1996) have also shown that because of shrinking all big coefficients towards zero, soft-thresholding ($T_{\tau,1}$) tends to have bigger bias and leads to less accurate results, although it is very stable. To overcome these disadvantages, recent research shows that L_p norm regulari-

zation generates more sparse solutions than L_1 norm regularization and more stable than L_0 norm regularization. It is possible to use the nonconvex L_p norm in Eq. 12, where $0 < p < 1$ (Chartrand 2007). The objective function $J(\mathbf{m})$ is written as

$$J(\mathbf{m}) = \|\tilde{\mathbf{D}} - \mathbf{F}\mathbf{m}\|_2^2 + \lambda \|\mathbf{C}^{-1}\mathbf{m}\|_p. \tag{32}$$

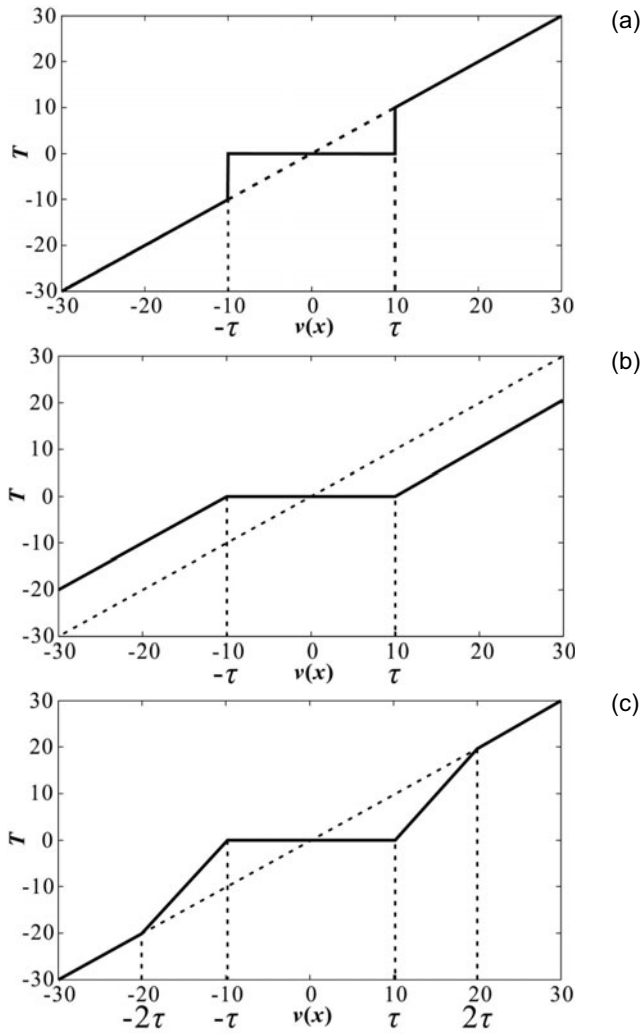


Fig. 2. Comparison of different thresholding functions: (a) hard-thresholding function ($\theta = 1/4$), (b) soft-thresholding function ($\theta \rightarrow \infty$), and (c) firm-thresholding function ($\theta = 1/2$).

In this case, we choose firm-thresholding introduced by Gao and Bruce (1997) as thresholding function. It is an intermediate between the soft and hard-thresholding, and can be formulated as

$$T_{\theta, 2\tau, p} = \begin{cases} v(\mathbf{m}_k), & \text{for } |v(\mathbf{m}_k)| > 4\theta\tau, \\ \frac{4\theta}{4\theta-1} [v(\mathbf{m}_k) - \tau \operatorname{sgn}(v(\mathbf{m}_k))], & \text{for } \tau < |v(\mathbf{m}_k)| \leq 4\theta\tau, \\ 0, & \text{for } |v(\mathbf{m}_k)| \leq \tau. \end{cases} \quad (33)$$

To demonstrate that firm-thresholding is an intermediate between soft- and hard-thresholding, we discuss the following cases.

For $\theta = 1/4$, firm-thresholding function becomes hard-thresholding function 29. Furthermore, letting $\theta \rightarrow \infty$, firm-thresholding function turns to soft-thresholding function 31. As a result, hard- and soft-thresholding are two extreme cases of firm-thresholding, which can be interpreted as an intermediate between soft and hard-thresholding. Figure 2 displays the shapes of hard-, soft-, and firm-thresholding functions.

4. DEBLENDING EXAMPLES

In order to confirm the effectiveness of the proposed method, three deblending examples are demonstrated in this section. We demonstrate two synthetic blended examples and a numerically blended example. Because of the instability of hard-thresholding, only the tests of ISTA and FISTA with soft- and firm-thresholding are shown and compared. All the data are deblended in the common receiver domain, and the curvelet transform is utilized as the sparse transform.

In addition, exponential-decreasing criterion is applied to control the thresholding value τ , which is easy to implement and can somewhat accelerate the convergence rate. And just as a slight clarification, if we use adequate number of iterations, the two approximation frames (ISTA and FISTA) would converge to the same optimal solution at different rates. Obviously, large number of iterations is not practical. With relative small number of iterations, FISTA converges to a better solution than ISTA.

As Chen *et al.* (2014a) did, we also use

$$SNR_{i,n} = 10 \times \log_{10} \frac{\|d_i\|_2^2}{\|d_i - d_{i,n}\|_2^2} \quad (34)$$

to evaluate the convergence performance and accuracy of the methods, where $SNR_{i,n}$ denotes signal-to-noise ratio (SNR) of the i -th source at the n -th iteration, $d_{i,n}$ represents the estimated model of the i -th source at the n -th iteration, and d_i stands for the true model for the i -th source.

4.1 Deblending simple synthetic blended data

This example contains three hyperbolic events without dip conflicts. The unblended and blended data are shown in Fig. 3.

We choose $\tau_{\max} = 0.25 \max \left(\mathbf{C}^{-1} \frac{2}{L(\nabla f(\mathbf{m}))} \mathbf{F}^T \tilde{\mathbf{D}} \right)$, $\tau_{\min} = 10^{-3}$, $\theta = 1/2$, and

the number of iterations is 40. We also add some random noise to the blended sections.

In Figure 4 panels (a) and (b) show the deblended results by using ISTA with soft- and firm-thresholding, and panels (c) and (d) show those by using FISTA with soft- and firm-thresholding. All these deblending results are nearly perfect.

The differences between the deblended and blended data are called blending noise shown in Fig. 5, where the spots are the reflections from other shots. It can be seen that most of the blending noise is removed by all the methods. The differences between the deblended and unblended data denoting the estimation error are illustrated in Fig. 6. It is seen from Fig. 6 that ISTA with soft-thresholding function has some obvious residuals, even though the residuals are very small. However, the estimation error sections of other three algorithms are almost zero.

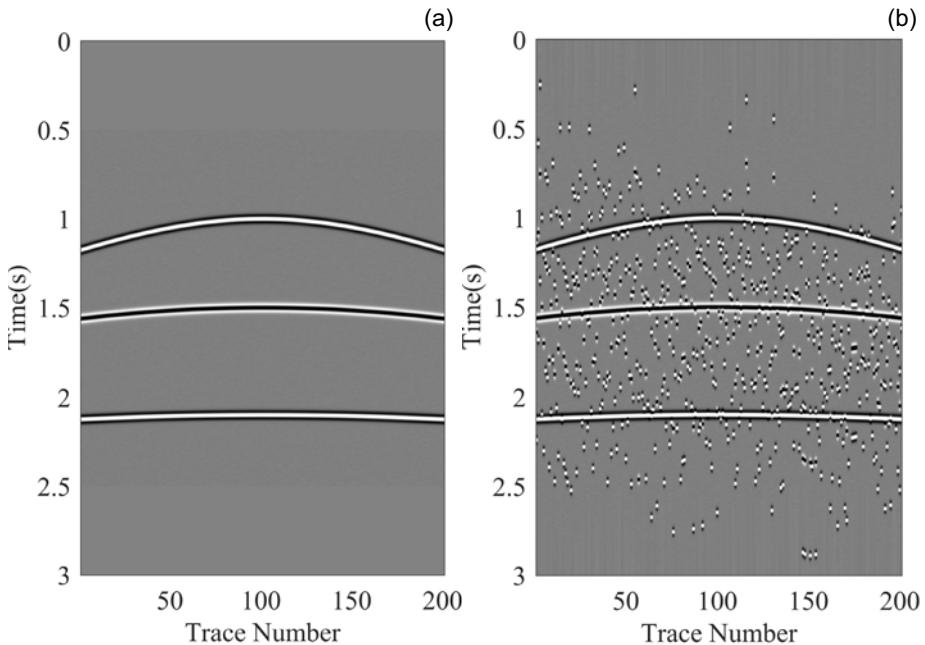


Fig. 3. Two numerical synthetic CRG datasets of the first example, unblended data (a), and blended data (b).

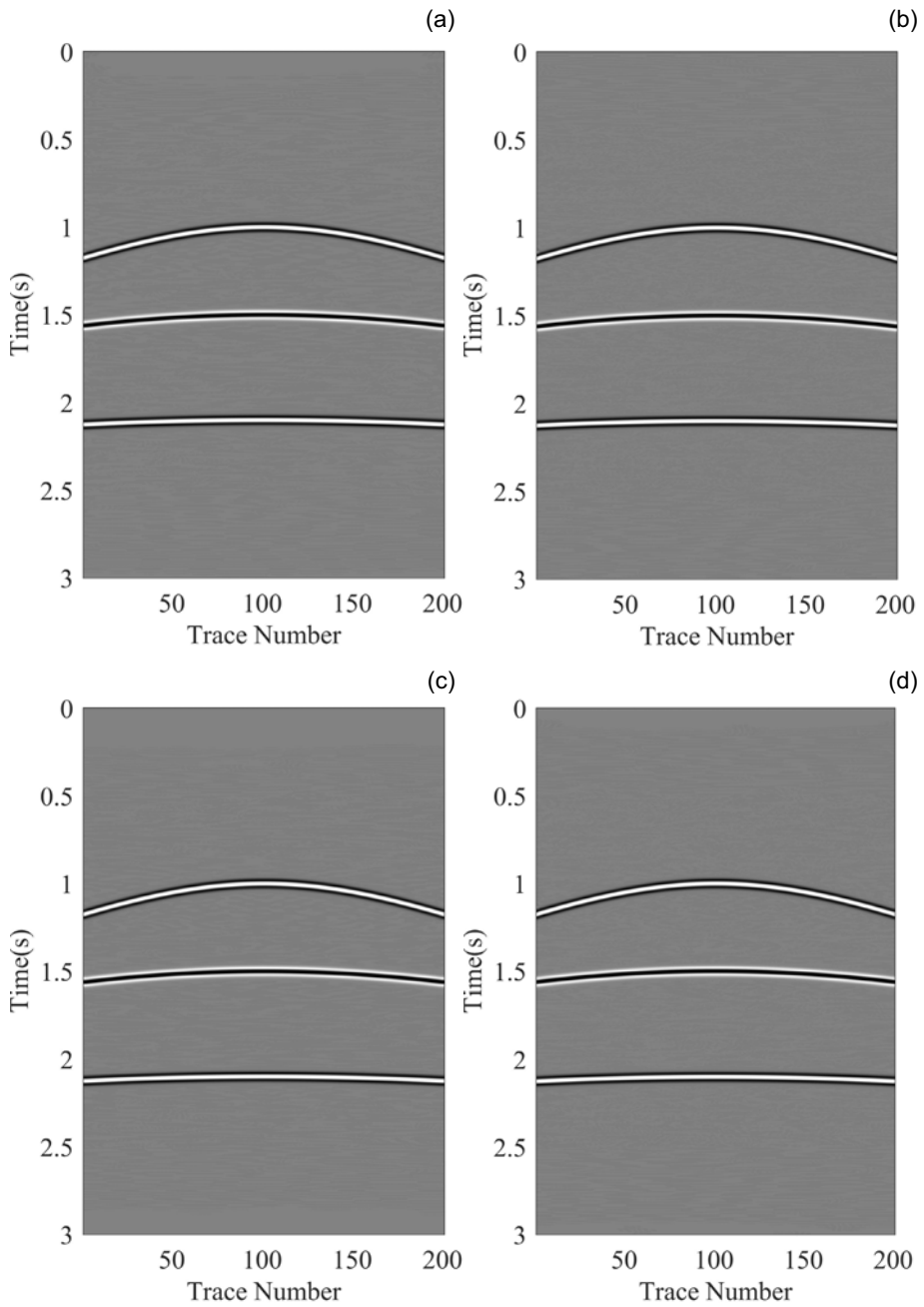


Fig. 4. Deblended result using ISTA with soft-thresholding (a), ISTA with firm-thresholding (b), FISTA with soft-thresholding (c), and FISTA with firm-thresholding (d).

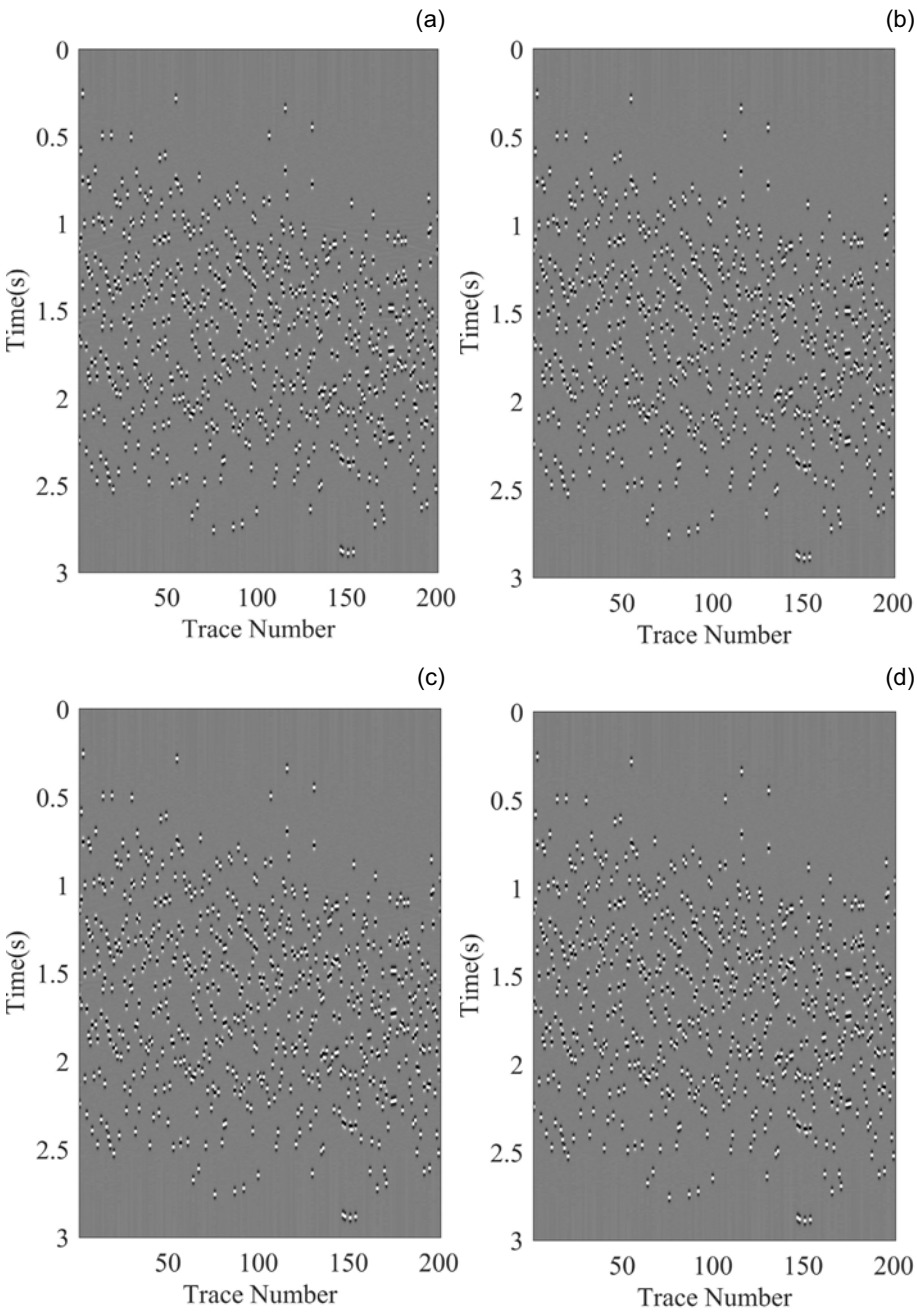


Fig. 5. Blending noise using ISTA with soft-thresholding (a), ISTA with firm-thresholding (b), FISTA with soft-thresholding (c), and FISTA with firm-thresholding (d).

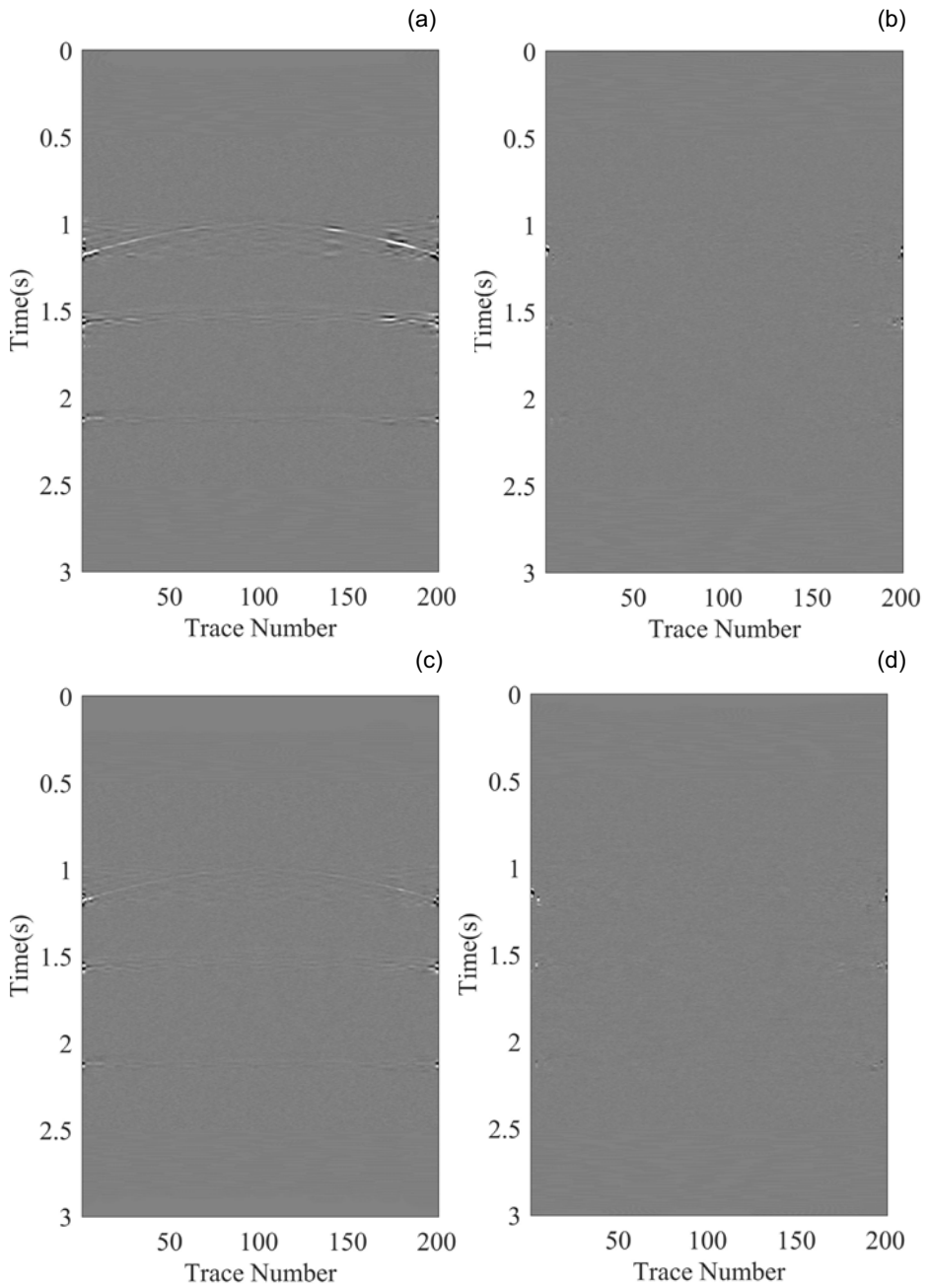


Fig. 6. Debrending estimation error using ISTA with soft-thresholding (a), ISTA with firm-thresholding (b), FISTA with soft-thresholding (c), and FISTA with firm-thresholding (d).

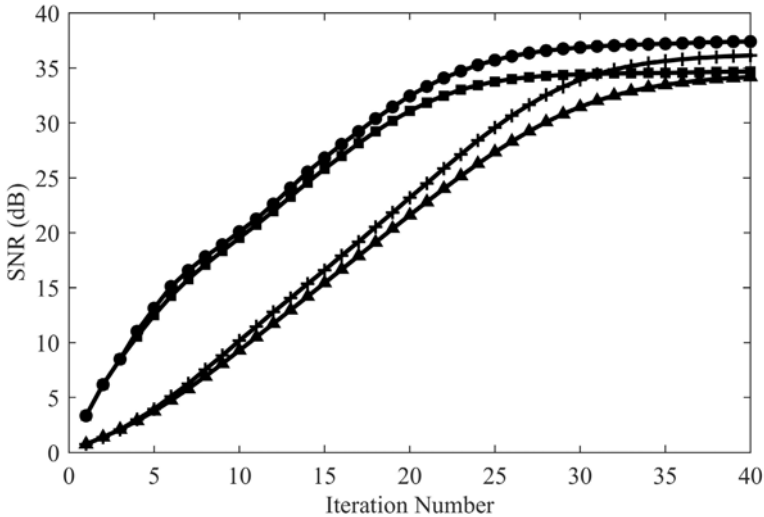


Fig. 7. Curves of SNR with iteration number corresponding to ISTA with soft-thresholding (▲), ISTA with firm-thresholding (■), FISTA with soft-thresholding (+), and FISTA with firm-thresholding (●).

The differences between the deblended and blended data are called blending noise shown in Fig. 5, where the spots are the reflections from other shots. It can be seen that most of the blending noise is removed by all the methods. The differences between the deblended and unblended data denoting the estimation error are illustrated in Fig. 6. It is seen from Fig. 6 that ISTA with soft-thresholding function has some obvious residuals, even though the residuals are very small. However, the estimation error sections of other three algorithms are almost zero.

The curves of convergence rate are shown in Fig. 7, in which FISTA behaves better than ISTA. Figure 7 also depicts that firm-thresholding is a superior choice of thresholding function than soft-thresholding.

4.2 Deblending complex synthetic blended data

The synthetic blended data whose geological model is shown in Fig. 8 contain much more events with many dip conflicts. These blended data are numerically calculated by using the staggered-grid finite-difference method. The original unblended and blended sections are shown in Fig. 9.

We choose $\tau_{\max} = 0.5 \max \left(\mathbf{C}^{-1} \frac{2}{L(\nabla f(\mathbf{m}))} \mathbf{F}^T \tilde{\mathbf{D}} \right)$, $\tau_{\min} = 10^{-1}$, $\theta = 1/2$, and

the number of iterations is 40. We add some random noise to the blended sections.

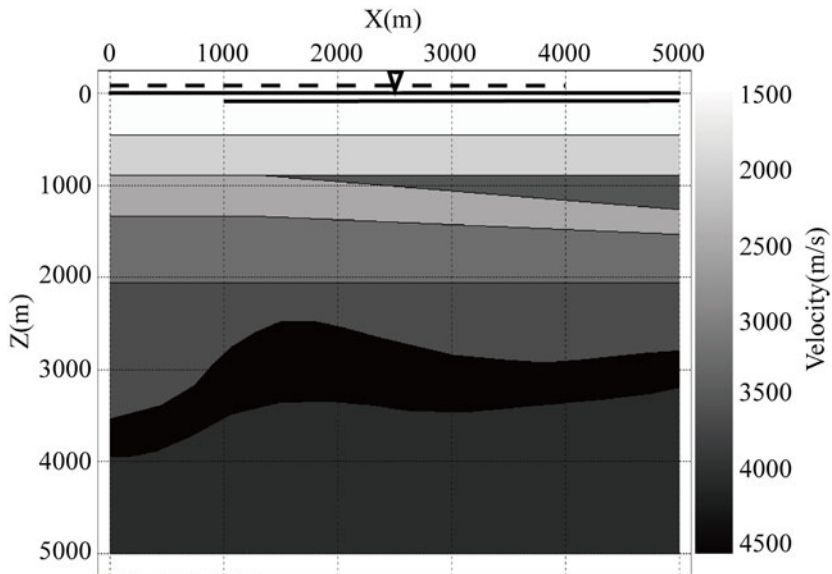


Fig. 8. Geological model used to calculate synthetic CRGs of the second example. ∇ stands for receiver position, the solid and dash line on the surface denote source lines 1 and 2, respectively.

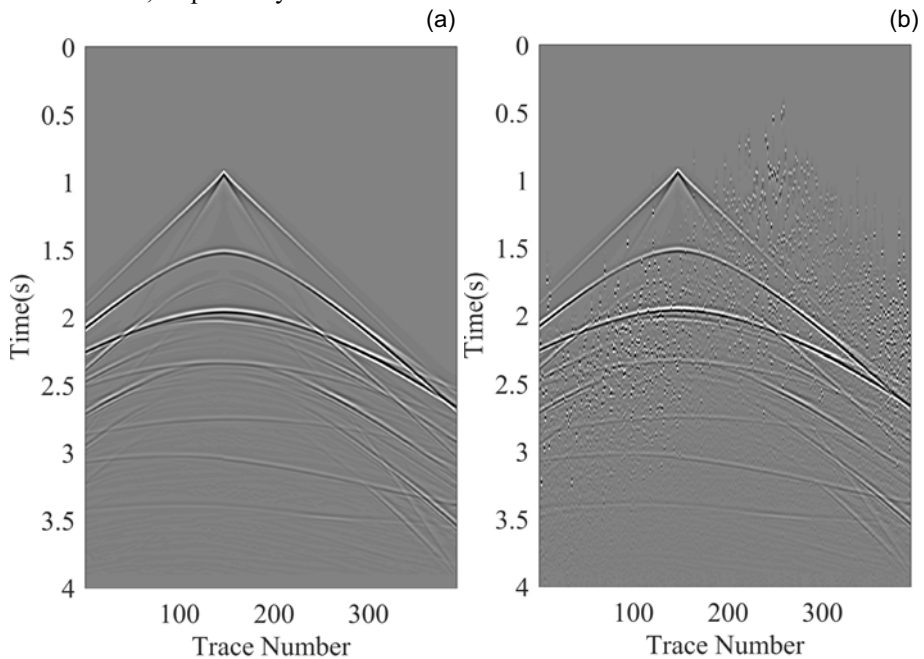


Fig. 9. Two synthetic CRG datasets of the second example: unblended data (a), and blended data (b).

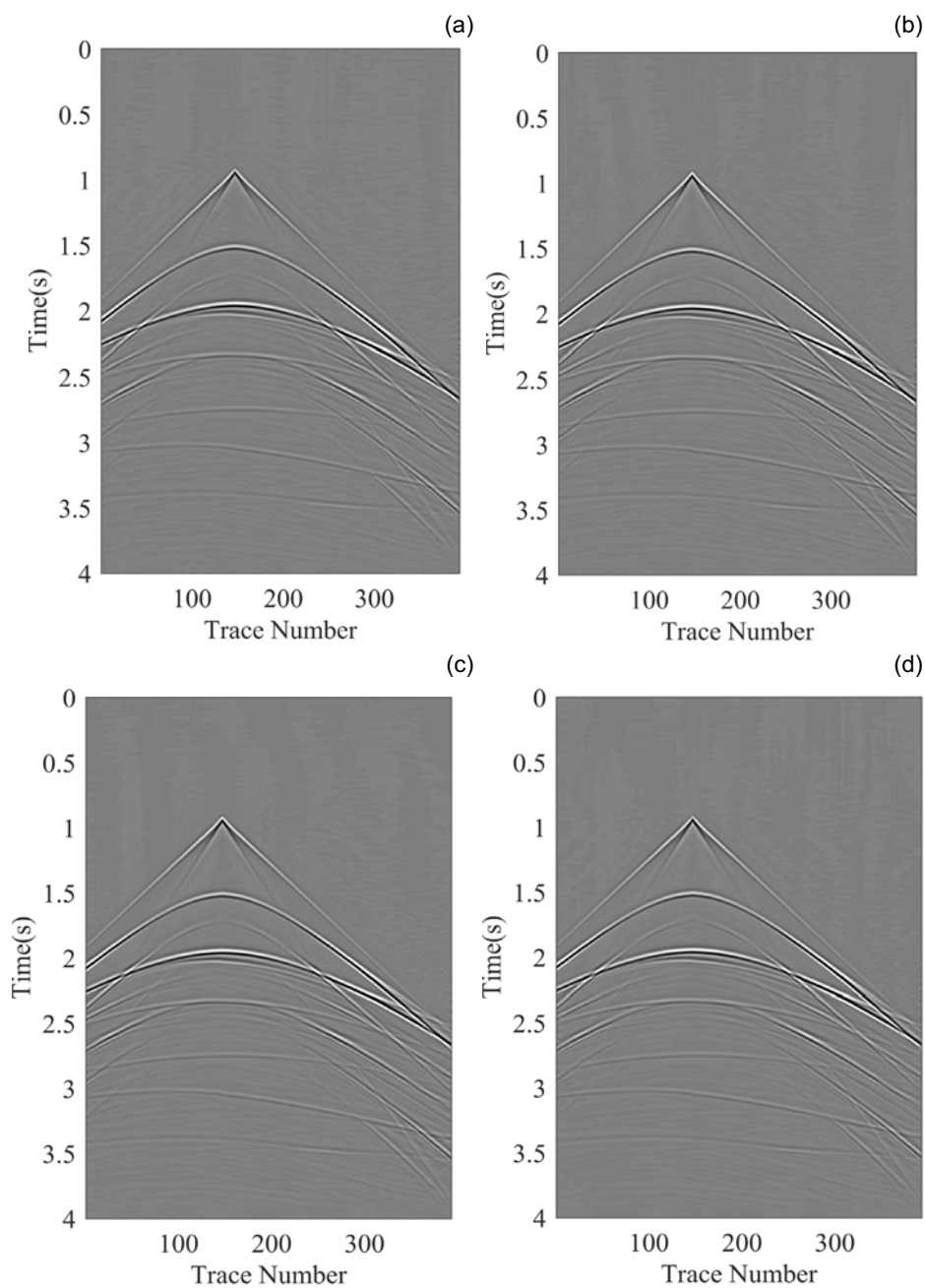


Fig. 10. Deblended result using ISTA with soft-thresholding (a), ISTA with firm-thresholding (b), FISTA with soft-thresholding (c), and FISTA with firm-thresholding (d).

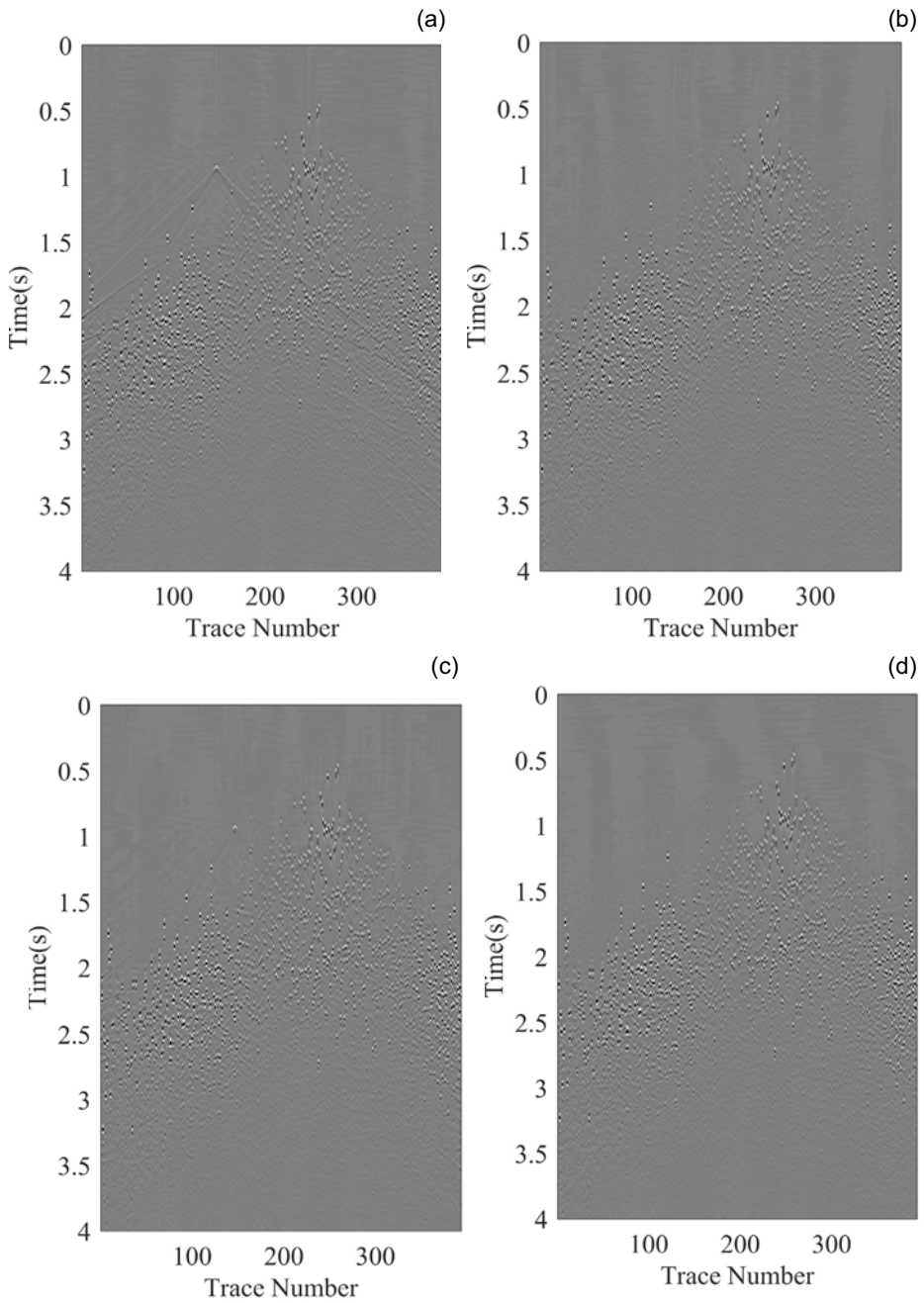


Fig. 11. Blending noise using ISTA with soft-thresholding (a), ISTA with firm-thresholding (b), FISTA with soft-thresholding (c), and FISTA with firm-thresholding (d).

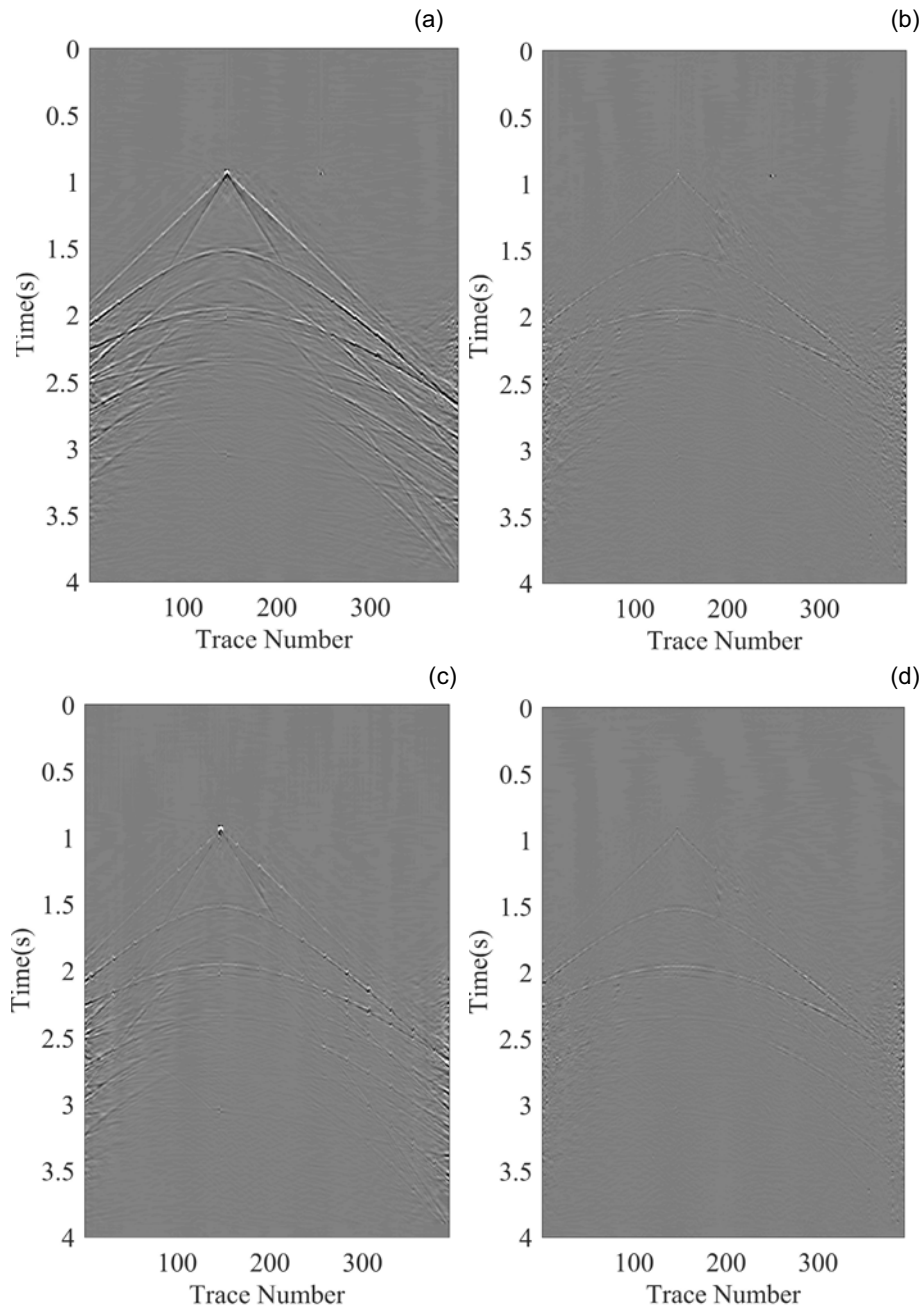


Fig. 12. Deblending estimation error using ISTA with soft-thresholding (a), ISTA with firm-thresholding (b), FISTA with soft-thresholding (c), and FISTA with firm-thresholding (d).

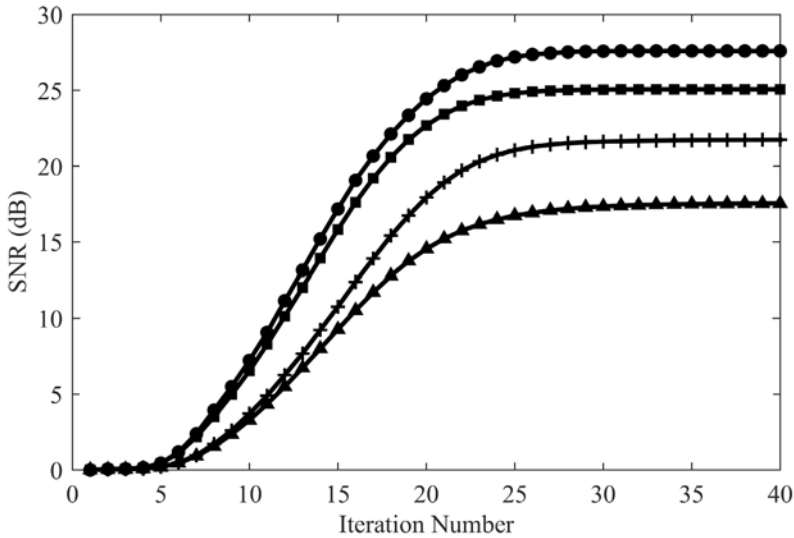


Fig. 13. Curves of SNR with iteration number corresponding to ISTA with soft-thresholding (\blacktriangle), ISTA with firm-thresholding (\blacksquare), FISTA with soft-thresholding (+), and FISTA with firm-thresholding (\bullet).

Figure 10 shows the deblending results, in which the useful reflections are very clear and all the results are acceptable. The blending noise sections in Fig. 11 show that most of the blending noise is attenuated by those methods. However, we can obviously find out the different effectiveness among those algorithms from the estimation error sections as shown in Fig. 12. The estimation error sections corresponding to ISTA and FISTA with soft-thresholding function are obviously worse than the sections corresponding to ISTA and FISTA with firm-thresholding. In addition, it can also be seen that when using the same thresholding functions, the results corresponding to ISTA have much more residuals than the results corresponding to FISTA.

Figure 13 shows the curves of SNR. It can be observed that FISTA with firm-thresholding is obviously better than the other methods, and it has the same behavior as the first example.

4.3 Deblending numerically blended field data

Common receiver gathers collected by OBC are used to generate two numerically blended field data. These blended data are separated to further test the effectiveness of our proposed method. From the previous tests, it can be seen that the effectiveness of FISTA with firm-thresholding is the best. As a result, we only separate the blended data by using FISTA with firm-thresholding.

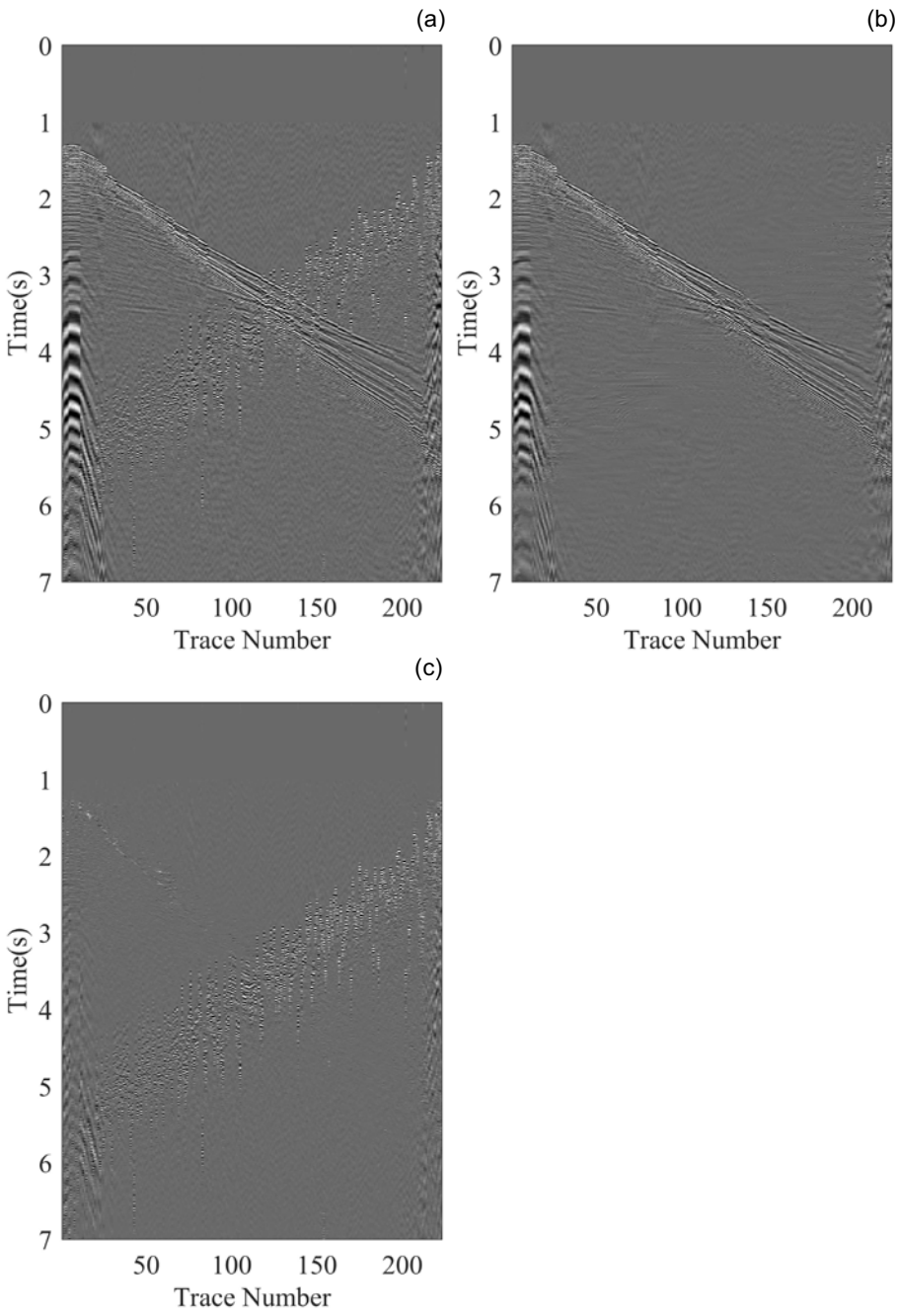


Fig. 14. Deblending example of field data of source line 1: blended gather (a), deblended result (b), and blending noise (c).

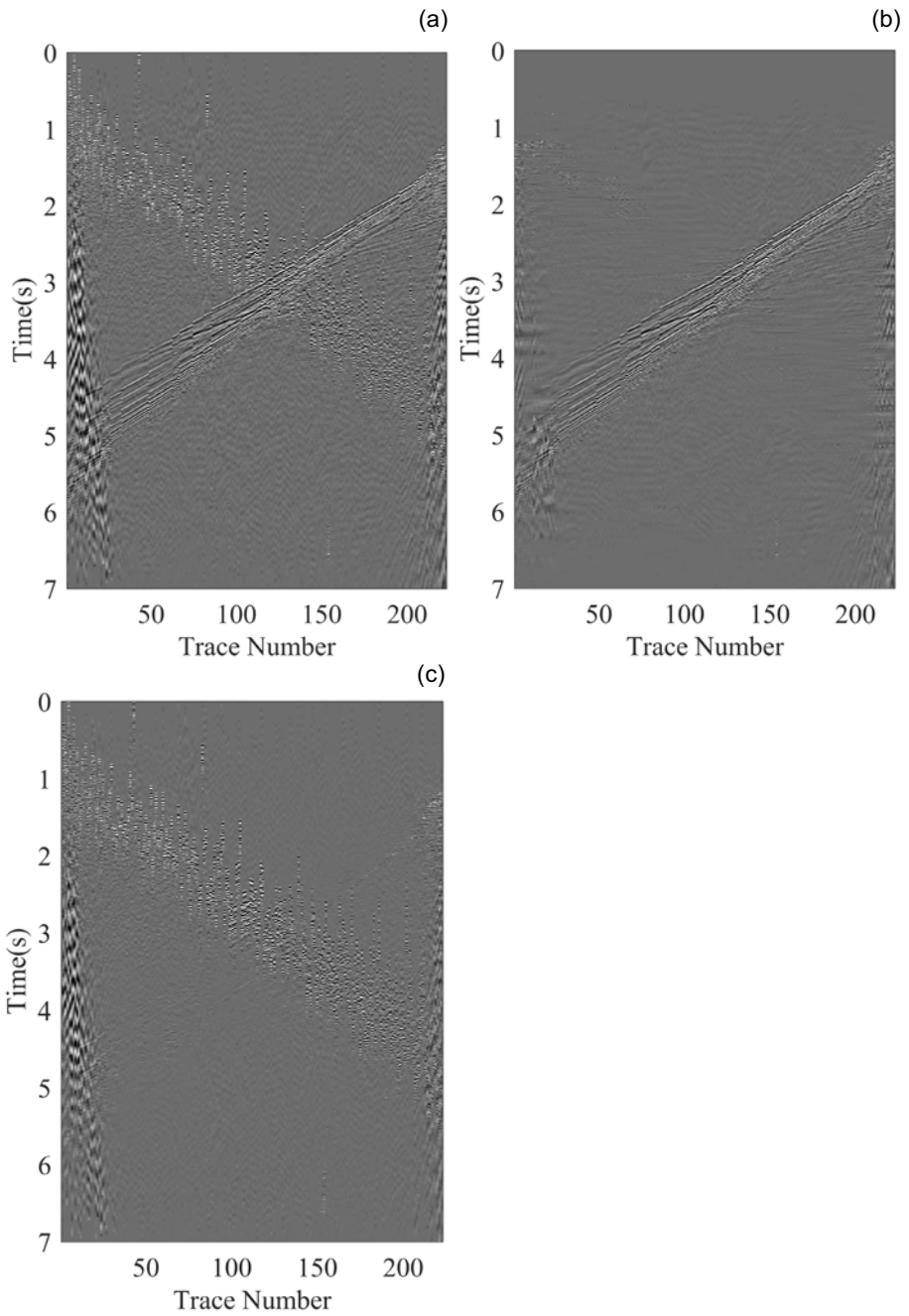


Fig. 15. Deblending example of field data of source line 2: blended gather (a), deblended result (b), and blending noise (c).

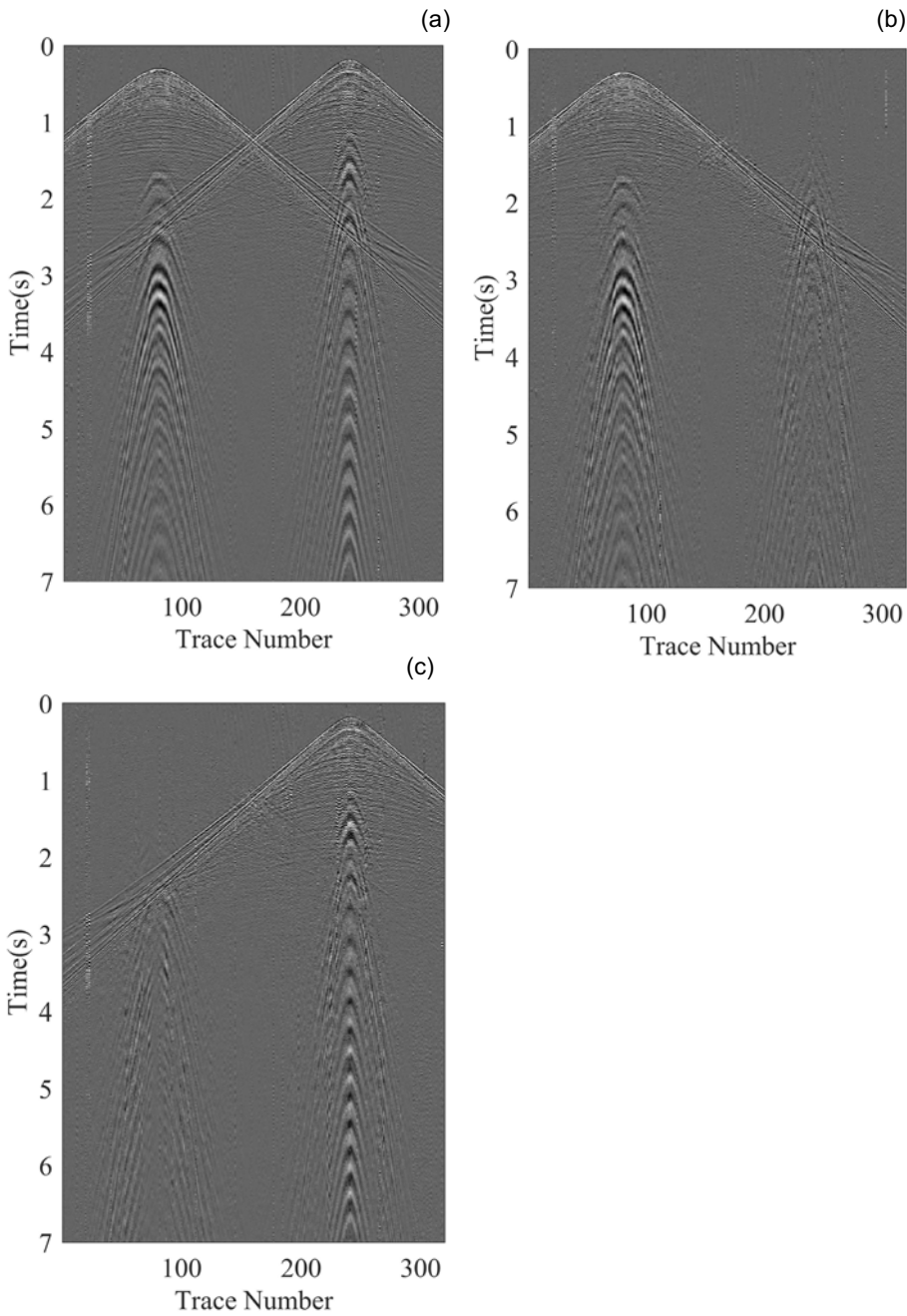


Fig. 16. Deblending example of field data of one common shot gather: blended gather (a), deblended result of shot 1 (b), and deblended result of shot 2 (c).

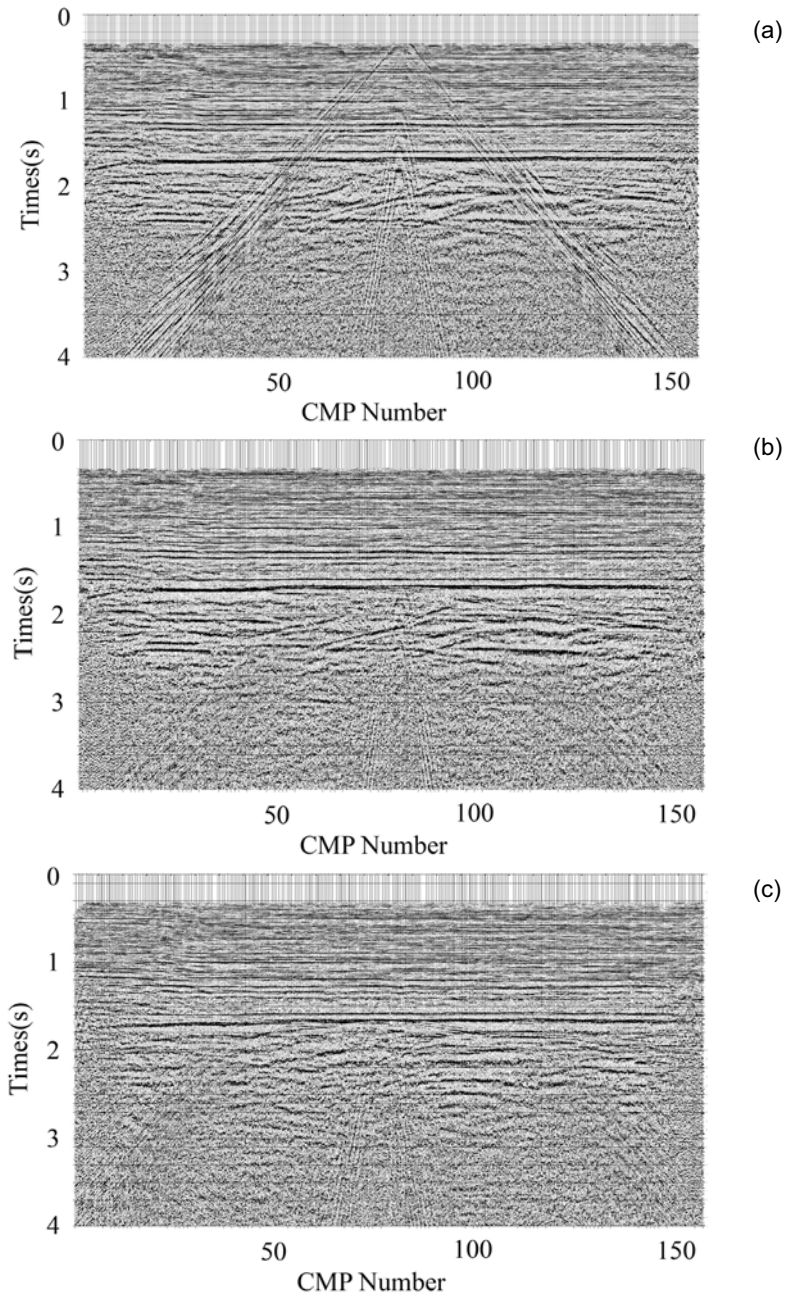


Fig. 17. Deblending example of field data of stack section: blended stack section (a), deblended stack section of source line 1 (b), and deblended stack section of source line 2 (c).

Two typical blended field datasets are shown in Figs. 14a and 15a, which are seriously contaminated by blending noise. After 30 iterations of FISTA with firm-thresholding, deblending results are shown in Figs. 14b and 15b, in which the reflections become much clear. The noise sections are shown in Figs. 14c and 15c, in which most of the blending noise is attenuated, although we can also see that some useful signals near the receiver point are damaged. The SNR increases from 2.2 to 11.6 dB for the first source, and from -2.2 to 9.2 dB for the second source. Although the deblending results are not so perfect, they are acceptable.

After processing all the common receiver gathers, we sort the data into common shot gather. One common shot gather is shown in Fig. 16. Panel (a) illustrates the blended profile, and panels (b) and (c) show deblended common shot gathers of the first and second shot, respectively. The results are likewise acceptable. Figure 17 displays the final stack sections of blended data, and separated data. We can see that the blending noise has been greatly decreased.

5. CONCLUSIONS

We have proposed an improved algorithm for deblending simultaneous-source data using FISTA with firm-thresholding. We have compared it with other alternative algorithms. All the algorithms, *i.e.*, ISTA with soft-thresholding, ISTA with firm-thresholding, FISTA with soft-thresholding and FISTA with firm-thresholding, have an acceptable effect to attenuate the blending noise; however, it can be concluded that in terms of approximation frame, FISTA has a faster convergence rate and higher accuracy than ISTA, and in terms of thresholding function, firm-thresholding function has a better convergence behavior and obtains a more accurate result.

When the blended seismic data become more complicated, it is more obvious that FISTA with firm-thresholding is the best choice to deblend the blended data. Moreover, our experiments indicate that it is possible to obtain accurate results within a small number of iterations using the proposed FISTA with firm-thresholding.

Acknowledgment. This work was supported in part by the 973 Program of China (2013CB228603), National Major Science and Technology Program (2011ZX05010-001, 2011ZX05024-001), National Natural Science Foundation of China (40974069, 41174119), and the Research of Novel Method and Technology of Geophysical Prospecting, CNPC (2014A-3609).

References

- Abma, R.L., T. Manning, M. Tanis, J. Yu, and M. Foster (2010), High quality separation of simultaneous sources by sparse inversion. **In: Proc. 72nd EAGE Conference and Exhibition incorporating SPE EUROPEC 2010**, DOI: 10.3997/2214-4609.201400611.
- Akerberg, P., G. Hampson, J. Rickett, H. Martin, and J. Cole (2008), Simultaneous source separation by sparse radon transform. **In: Proc. SEG Annual Meeting, 9-14 November 2008, Las Vegas, USA**, SEG-2008-2801, Society of Exploration Geophysicists.
- Beasley, C.J. (2008), A new look at marine simultaneous sources, *The Leading Edge* **27**, 7, 914-917, DOI: 10.1190/1.2954033.
- Beck, A., and M. Teboulle (2009), A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.* **2**, 1, 183-202, DOI: 10.1137/080716542.
- Bruce, A.G., and H.-Y. Gao (1996), Understanding waveshrink: variance and bias estimation, *Biometrika* **83**, 4, 727-745, DOI: 10.1093/biomet/83.4.727.
- Candes, E.J., and L. Demanet (2005), The curvelet representation of wave propagators is optimally sparse, *Commun. Pure Appl. Math.* **58**, 11, 1472-1528, DOI: 10.1002/cpa.20078.
- Candes, E.J., and T. Tao (2005), Decoding by linear programming, *IEEE Trans. Inf. Theory* **51**, 12, 4203-4215, DOI: 10.1109/TIT.2005.858979.
- Chambolle, A. (2004), An algorithm for total variation minimization and applications, *J. Math. Imaging Vis.* **20**, 1-2, 89-97, DOI: 10.1023/B:JMIV.0000011325.36760.1e.
- Chartrand, R. (2007), Exact reconstruction of sparse signals via nonconvex minimization, *IEEE Signal Process. Lett.* **14**, 10, 707-710, DOI: 10.1109/LSP.2007.898300.
- Chen, Y., and J. Ma (2014), Random noise attenuation by f-x empirical-mode decomposition predictive filtering, *Geophysics* **79**, 3, V81-V91, DOI: 10.1190/geo2013-0080.1.
- Chen, Y., S. Fomel, and J. Hu (2014a), Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization, *Geophysics* **79**, 5, V179-V189, DOI: 10.1190/geo2013-0449.1.
- Chen, Y., J. Yuan, Z. Jin, K. Chen, and L. Zhang (2014b), Deblending using normal moveout and median filtering in common-midpoint gathers, *J. Geophys. Eng.* **11**, 4, 045012, DOI: 10.1088/1742-2132/11/4/045012.
- Daubechies, I., M. Defrise, and C. De Mol (2004), An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Commun. Pure Appl. Math.* **57**, 11, 1413-1457, DOI: 10.1002/cpa.20042.

- Donoho, D.L. (2006), For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution, *Commun. Pure Appl. Math.* **59**, 6, 797-829, DOI: 10.1002/cpa.20132.
- Doulgeris, P., K. Bube, G. Hampson, and G. Blacquiere (2012), Convergence analysis of a coherency-constrained inversion for the separation of blended data, *Geophys. Prospect.* **60**, 4, 769-781, DOI: 10.1111/j.1365-2478.2012.01088.x.
- Figueiredo, M.A., and R.D. Nowak (2003), An EM algorithm for wavelet-based image restoration, *IEEE Trans. Image Process.* **12**, 8, 906-916, DOI: 10.1109/TIP.2003.814255.
- Gao, H.-Y., and A.G. Bruce (1997), Waveshrink with firm shrinkage, *Statistica Sinica* **7**, 4, 855-874.
- Huo, S., Y. Luo, and P. Kelamis (2009), Simultaneous sources separation via multi-directional vector-median filter. **In:** *Proc. SEG Annual Meeting, 25-30 October 2009, Houston, USA*, SEG-2009-0031, Society of Exploration Geophysicists.
- Lin, T.T.Y., and F.J. Herrmann (2009), Designing simultaneous acquisitions with compressive sensing. **In:** *Proc. 71st EAGE Conference and Exhibition incorporating SPE EUROPEC 2009*, DOI: 10.3997/2214-4609.201400276.
- Mahdad, A., P. Doulgeris, and G. Blacquiere (2011), Separation of blended data by iterative estimation and subtraction of blending interference noise, *Geophysics* **76**, 3, Q9-Q17, DOI: 10.1190/1.3556597.
- Mahdad, A., P. Doulgeris, and G. Blacquiere (2012), Iterative method for the separation of blended seismic data: discussion on the algorithmic aspects, *Geophys. Prospect.* **60**, 4, 782-801, DOI: 10.1111/j.1365-2478.2012.01084.x.
- Qu, S., H. Zhou, H. Chen, S. Zu, and L. Zhi (2014), Separation of simultaneous vibroseis data. **In:** *Proc. SEG Annual Meeting, 26-31 October 2014, Denver, USA*, SEG-2014-0182, Society of Exploration Geophysicists.

Received 9 December 2014

Received in revised form 29 May 2015

Accepted 14 July 2015